

09/629,831

REMARKS

Claims 1-17, all the claims pending in the application, stand rejected on prior art grounds. Applicants respectfully traverse these rejections based on the following discussion.

L The Prior Art Rejections

Claims 1, 6, and 11 stand rejected under 35 U.S.C. §103(a) as being unpatentable over Kostoff et al., hereinafter "Kostoff" (U.S. Patent No. 5,440,481 in view of Kirsch et al., hereinafter "Kirsch" (U.S. Patent No. 6,070,158). Claims 2-5, 7-10, and 12-17 stand rejected under 35 U.S.C. §103(a) as being unpatentable over Kostoff and Kirsch, and further in view of Kobayashi (U.S. Patent No. 5,742,834) and Turney (U.S. Patent No. 6,470,307). Applicants respectfully traverse these rejections based on the following discussion.

A. The Rejection Based on Kostoff in view of Kirsch

As explained on page 4, lines 4-9 of the application, the invention allows the user to specify the size of the vector space model to be used in text clustering of a document corpus, as well as the maximum number of words that can occur in a phrase. The invention will find all of the phrases, up to the user specified length, that occur with the greatest frequency. The total number of phrases returned will depend upon the user specified maximum dictionary size.

One distinction of the invention when compared to Kostoff is that the invention avoids maintaining a list of all potential phrases in the text corpus. The problem with maintaining all potential phrases is that the number of phrases grows exponentially with the size of the corpus. The invention avoids this problem by fixing the size of the dictionary up front (user specified maximum dictionary size, M), then finding the M most frequent words and then only creating phrases using these M most frequent words.

09/629,831

This limits the number of phrases that have to be considered, to at most $M*(M-1)*(M-2)...(M-N-1)$ (where N is the phrase length), so that the computational memory requirements are independent of the document corpus size. The invention can then specify a dictionary size that makes the number of phrases that have to be kept track of tractable.

To the contrary, the Kostoff patent creates a list of all words and N -word phrases sorted by frequency. This is not practical for a large text corpus since such a list would be too large for most computer memory to hold. The Hashtable mentioned in Kirsch is not germane to this invention because it is created as an index over an existing word/phrase dictionary, and the terminology has been removed from the claims to avoid any confusion. The claimed invention deals with dictionary creation, which is a prior step to index creation.

Therefore, it is applicants' position that the combination of Kostoff and Kirsch does not teach or suggest "creating a dictionary of most frequently occurring words in said documents as limited by said maximum dictionary size; determining a frequency of phrases in each of said documents that contain only words in said dictionary; adding most frequently occurring phrases to said dictionary; and outputting said most frequently occurring words and said most frequently occurring phrases as said dictionary" as defined by independent claims 1 and 11 and similarly defined by independent claim 6. Previous methodologies that have suggested a lexical phrase generation technique have not described the space and time efficient implementation for discovering such phrases that the invention utilizes. The invention's implementation is designed to quickly find a maximal frequency term dictionary of a given size using the smallest possible amount of memory.

Therefore, because the proposed combination of references does not teach or suggest the claimed invention, Applicants respectfully submit that independent claims 1, 6, and 11 are patentable over the prior art of record. In view the foregoing, the Examiner is respectfully requested to reconsider and withdraw this rejection.

09/629,831

**B. The Rejection Based on Kostoff in view of Kirsch
and further in view of Kobayashi and Turney**

With respect to dependent claims 2-5, 7-10, and 12-17, the Office Action makes reference to the prior art Kobayashi and Turney as teaching concepts such as removing punctuation, replacing words with synonyms, removing stop words, removing duplicate words, clustering, etc. Therefore, the additional prior art references are not utilized to teach or suggest (and do not teach or suggest) the claimed features defined by independent claims 1, 6, and 11. Therefore, it is Applicant's position that the proposed combination of all references still does not teach or suggest "creating a dictionary of most frequently occurring words in said documents as limited by said maximum dictionary size; determining a frequency of phrases in each of said documents that contain only words in said dictionary; adding most frequently occurring phrases to said dictionary; and outputting said most frequently occurring words and said most frequently occurring phrases as said dictionary" as defined by independent claims 1 and 11 and similarly defined by independent claim 6. Therefore, it is Applicant's position that none of the prior art of record teaches or suggests the invention defined by independent claims 1, 6, and 11 and that such independent claims are patentable over the prior art record.

Further, dependent claims 2-5, 7-10, and 12-17 are similarly patentable, not only by virtue of their dependency from a patentable independent claim, but also by virtue of the additional features of the invention they define. Therefore, Applicants submit that dependent claims 2-5, 7-10, and 12-17 are patentable over the prior art of record and respectfully request that the Examiner reconsider and withdraw this rejection.

09/629,831

II. Formal Matters and Conclusion

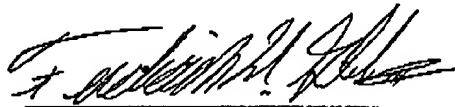
In view of the foregoing, Applicants submit that claims 1-17, all the claims presently pending in the application, are patentably distinct from the prior art of record and are in condition for allowance. The Examiner is respectfully requested to pass the above application to issue at the earliest possible time.

Should the Examiner find the application to be other than in condition for allowance, the Examiner is requested to contact the undersigned at the local telephone number listed below to discuss any other changes deemed necessary.

Please charge any deficiencies and credit any overpayments to Attorney's Deposit Account Number 09-0441.

Respectfully submitted,

Dated: 4/13/04



Frederick W. Gibb, III
Reg. No. 37,629

McGinn & Gibb, P.C.
2568-A Riva Road
Suite 304
Annapolis, MD 21401
301-261-8071
Customer Number: 28211